The PAICE Suite: Using Extended Harmonic Oscillators to Identify and Understand Circadian Rhythms in Large Datasets

Hannah De los Santos[1], Emily J. Collins[2], Catherine F. Mann[2], Meaghan S. Jankowski[2], April W. Sagan[3], Kristin P. Bennett[1,3], Jennifer M. Hurley[2]

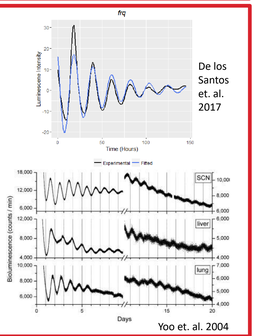
[1] Dept. of Computer Science, Rensselaer Polytechnic Institute, Troy, NY, USA, [2] Dept. of Biological Science, Rensselaer Polytechnic Institute, Troy, NY, USA,
[3] Dept. of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY, USA

1. Abstract

We present the PAICE (Pipeline for Amplitude Integration of Circadian Exploration) Suite, a group of tools to detect and understand circadian rhythms in large datasets. Circadian rhythms are endogenous cycles of approximately 24 hours reinforced by external cues such as light. These cycles are typically modeled as harmonic oscillators with fixed amplitude. Using experimental time series data, we demonstrate that many circadian genes exhibit non-harmonic oscillations (decreasing or increasing amplitude). By fitting Extended Harmonic Circadian Oscillation (ECHO) models which include an amplitude change (AC) coefficient, we detected additional circadian genes that were not identified by the current standards. Unlike these standards, in our synthetic validation datasets, ECHO maintains high accuracy and phase recall despite increases in noise and resolution for all AC coefficient categories. We then built the ECHO functionality into a freely available, easy-to-use interface for circadian biologists with two sections: finding rhythms in uploaded data, and visualizing these results, available at www.github.com/delosh653/ECHO. Further, we leverage these new AC categories to create the second application in the PAICE Suite, the ECHO Native Circadian Ontological Rhythmicity Explorer (ENCORE), available at www.github.com/delosh653/ENCORE. ENCORE jointly performs gene set enrichment and analysis of protein-protein interactions between AC categories, allowing users to fully derive biological understanding of the function of these groups and their interactions. For biologists seeking understanding of rhythms in their data, this application is easy to navigate and generates publication-worthy images, providing a welcome enhancement to the overwhelming output given by standard gene enrichment sites. Utilizing the PAICE Suite, we were able to discover and understand novel rhythms in datasets using common model organisms. Not only did AC categories reflect experimental conditions, but they also corresponded to separate functional gene mechanisms, meaning that the measurement of AC categories is vital to rhythmic gene understanding.

2. Background

Circadian rhythms are ~24 hour rhythms that often follow oscillatory expression patterns. Therefore, common existing methodologies, such as JTK_CYCLE, compare to reference cosine curves. These models, however, are fixed amplitude, which means that they're not designed to handle damping rhythms, commonly seen *in vivo*. Thus, fixed amplitude models could miss genes with these damped patterns. While the underlying cause of this damping is still debated, the prevalence of this damping is difficult to overstate.



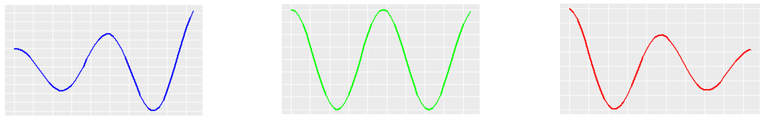
3. The ECHO Model

In order to capture this damping, we developed the first part of the PAICE Suite: the ECHO model, which uses extended harmonic oscillators to capture amplitude change:

$$x(t) = Ae^{\frac{-\gamma t}{2}}\cos(\omega t + \varphi) + y$$

Most of the parameters stay the same from the traditional fixed amplitude model, such as amplitude and frequency, but what we care about most here is the amplitude change (AC) coefficient, γ, which captures the amount of damping in the system.

4. Categorizing Rhythms By AC Coefficient Values



Forced **Harmonic** **Damped**

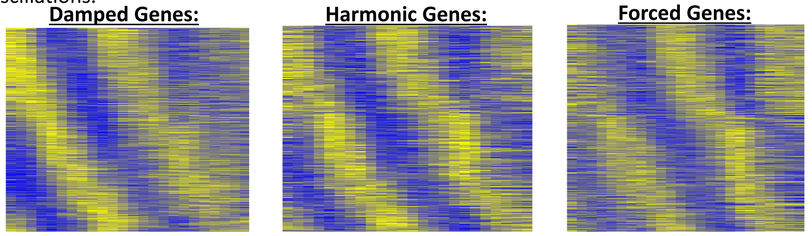
Using the AC coefficient, we can divide these rhythms into categories. This is automatically done through the ECHO[1] application, which provides efficient ECHO calculations on any rhythmic dataset, as well as automatic visualizations, including heat maps and gene expression plots.

[1]github.com/delosh653/ECHO

5. ECHO identifies circadian rhythms of all categories.

When we analyzed previously published large-scale data sets using ECHO, (below is a representative data set from *Neurospora crassa* (Hurley et al., 2018)), we found that each of the data sets that we investigated contained large numbers of non-harmonic oscillations.

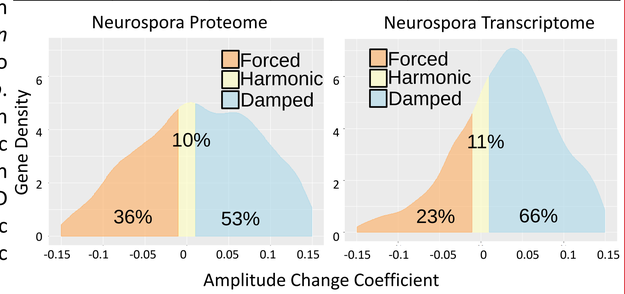
Damped Genes: **Harmonic Genes:** **Forced Genes:**



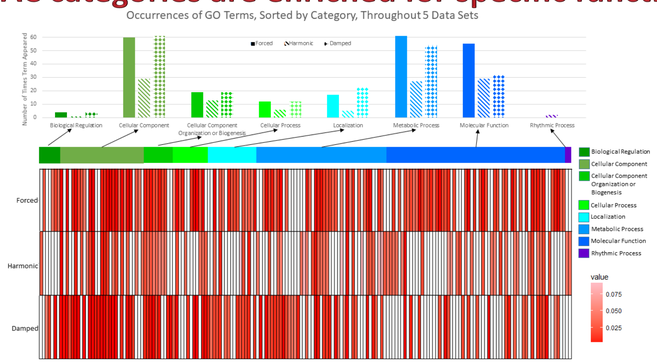
6. The conditions of sampling and the level of output that is being sampled determine the damped to forced gene ratio.

We analyzed the ratio of AC categories and related these ratios to the conditions under which the samples were acquired. We found an increase in damped genes in samples acquired *in vitro* as compared to those sampled *in vivo*. We also noted an increase in harmonic genes between samples taken in LD and DD and proteomic vs transcriptomic sampling.

Sampling	*In Vivo*			*In Vitro*		
Organism	*Mus musculus* (Liver)	*Anopheles gambiae* (Heads)		*Neurospora crassa*		*Mus musculus* (NIH3T3 Fibroblasts)
Zeitgeber scheme	DD		LD	DD		Forskolin synchronized
Technique	Microarray			RNA-seq	TMT-MS	Microarray
% Damped	37.5	48.5	34.7	66.2	53.0	73.1
% Forced	50.7	37.2	42.4	23.3	36.8	21.2
% Harmonic	11.8	14.4	22.9	10.6	10.2	5.7

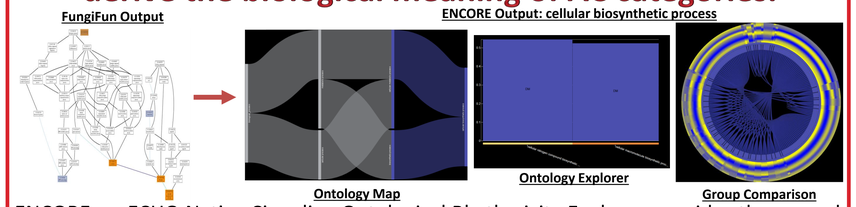


7. All AC categories are enriched for specific functions.



Gene ontologies (GOs) associated with metabolic functions were specific to Forced genes. GOs associated with the regulation of transcription/translation were specific to Damped genes. GOs associated with Harmonic genes were specific to cellular rhythms. o really dig into the function of these AC categories, we utilize the second part of the PAICE Suite: ENCORE.

8. ENCORE connects GO and STRING information to derive the biological meaning of AC categories.

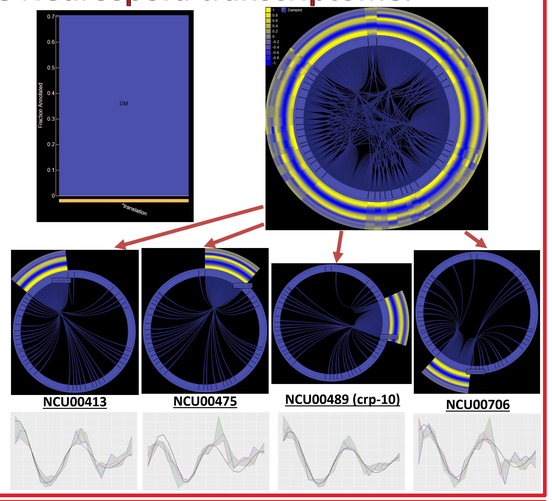


ENCORE, or ECHO Native Circadian Ontological Rhythmicity Explorer, provides the second part of PAICE. ENCORE[2] connects the disparate datasets of gene ontology information and STRING protein-protein interaction networks, creating the first rhythmicity specific interface to analyze the connections and biological meaning of AC categories. This interactive app provides GO data in an easily digestible format, providing a welcome enhancement to the overwhelming output given by standard gene ontology sites.

[2]github.com/delosh653/ENCORE

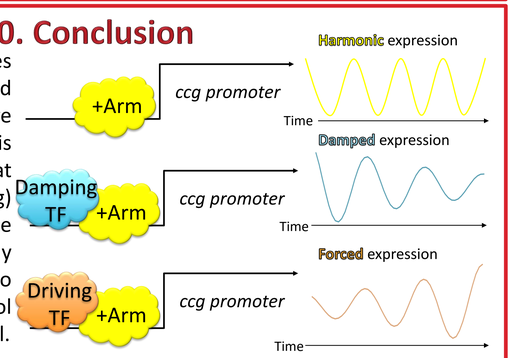
9. ENCORE confirms specificity of translation of damped genes in the Neurospora transcriptome.

Using ENCORE, we honed in on translation, which we found was significantly enriched for the damped category, with .705 of the total amount of genes related to translation appearing in our group and a fold enrichment of 1.474. In looking at the top 150 protein-protein connections for damped genes related to protein, we find 4 highly-rhythmic and connected genes: NCU00706, NCU00413, NCU00475, and NCU00489 (crp-10), each pertaining to ribosomal proteins.



10. Conclusion

Using PAICE, our data suggests genes that fall into damped, harmonic, and forced oscillatory categories have distinct biological roles. This indicates that the regulation that underlies the damping (or driving) process is not an artifact of the sampling process but biologically relevant. Further research to determine the mechanistic control underlying circadian damping is vital.



11. Acknowledgements

Funding was provided by an NIH (NIBIB U01 EB02246) grant, a DOE (PNNL 47818) grant, and RPI startup funds to J. Hurley, an RPI Presidential Fellowship to H. De Los Santos, an NSF (#1331023) grant to K. Bennett and an NIH T32 training grant to E. Collins.